

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-252268

(43) 公開日 平成9年(1997)9月22日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 B 3/23			H 0 4 B 3/23	
G 1 0 L 3/00	5 5 1		G 1 0 L 3/00	5 5 1 A
	3 0 1			3 0 1 D
H 0 4 M 1/00			H 0 4 M 1/00	H

審査請求 有 請求項の数11 O L (全 13 頁)

(21) 出願番号 特願平8-58633

(22) 出願日 平成8年(1996)3月15日

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 高木 啓三郎

東京都港区芝五丁目7番1号 日本電気株式会社内

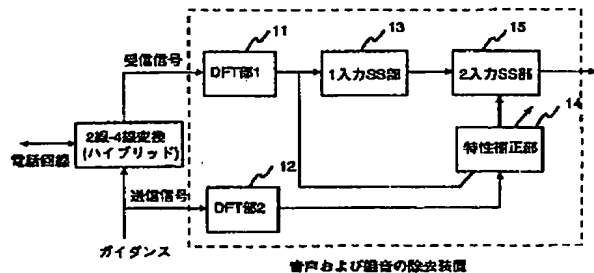
(74) 代理人 弁理士 京本 直樹 (外2名)

(54) 【発明の名称】 音声および雑音の除去装置、音声認識装置

(57) 【要約】

【課題】 自己の送出したガイダンス音声だけでなく相手話者の周囲の雑音をも除去し、高性能な音声および雑音の除去装置と、音声認識装置を提供することにある。

【解決手段】 電話の受信信号をスペクトルに変換する第1のDFT部11と、電話の送信信号をスペクトルに変換する第2のDFT部12と、第1のDFT部11の出力から相手の話者の音声の周囲雑音を除去する1入力スペクトルサブトラクション部13と、第1のDFT部11の出力と第2のDFT部12の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、第2のDFT部12の出力信号を受信側の特性に合わせて補正する特性補正部14と、1入力スペクトルサブトラクション部13の出力から特性補正部14の出力を差し引くことにより2入力のスペクトルサブトラクションを行なう2入力スペクトルサブトラクション部15とで構成される。



**【特許請求の範囲】**

【請求項 1】電話の受信信号をスペクトルに変換する第 1 の D F T 部と、  
電話の送信信号をスペクトルに変換する第 2 の D F T 部と、  
前記第 1 の D F T 部の出力から相手の話者の音声の周囲雑音を除去する 1 入力スペクトルサブトラクション部と、  
前記第 1 の D F T 部の出力と前記第 2 の D F T 部の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正する特性補正部と、  
前記 1 入力スペクトルサブトラクション部の出力から前記特性補正部の出力を差し引くことにより 2 入力のスペクトルサブトラクションを行なう 2 入力スペクトルサブトラクション部とで構成される音声および雑音の除去装置。

【請求項 2】電話の受信信号をスペクトルに変換する第 1 の D F T 部と、  
電話の送信信号をスペクトルに変換する第 2 の D F T 部と、  
前記第 1 の D F T 部の出力から相手の話者の音声の周囲雑音を除去する 1 入力スペクトルサブトラクション部と、  
前記 1 入力スペクトルサブトラクション部の出力と前記第 2 の D F T 部の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正する特性補正部と、  
前記 1 入力スペクトルサブトラクション部の出力から前記特性補正部の出力を差し引くことにより 2 入力のスペクトルサブトラクションを行なう 2 入力スペクトルサブトラクション部とで構成される音声および雑音の除去装置。

【請求項 3】前記特性補正部が電話の送信信号の存在する区間を用いて送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正することを特徴とする請求項 1 または 2 記載の音声および雑音の除去装置。

【請求項 4】前記 2 入力スペクトルサブトラクション部の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部を備え、  
前記 1 入力スペクトルサブトラクション部は前記音声検出部が音声区間検出中でない区間から除去する雑音を求めることを特徴とする請求項 1 または 2 記載の音声および雑音の除去装置。

【請求項 5】前記 1 入力スペクトルサブトラクション部は前記音声検出部が音声区間検出中でない区間かつ電話の送信信号の存在しない区間から除去する雑音を求めることを特徴とする請求項 4 記載の音声および雑音の除去

装置。

【請求項 6】前記特性補正部は前記音声検出部が音声区間検出中でない区間から電話の送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正することを特徴とする請求項 4 または 5 記載の音声および雑音の除去装置。

【請求項 7】前記特性補正部は音声検出部が音声区間検出中でない区間かつ電話の送信信号の存在する区間から電話の送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正することを特徴とする請求項 6 記載の音声および雑音の除去装置。

【請求項 8】前記 2 入力スペクトルサブトラクション部の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部を備え、

前記特性補正部が前記音声検出部が音声区間検出中でない区間かつ電話の送信信号の存在する区間を用いて送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正することを特徴とする請求項 3 記載の音声および雑音の除去装置。

【請求項 9】電話の受信信号をスペクトルに変換する第 1 の D F T 部と、

電話の送信信号をスペクトルに変換する第 2 の D F T 部と、

前記第 1 の D F T 部の出力から相手の話者の音声の周囲雑音を除去する 1 入力スペクトルサブトラクション部と、

前記第 1 の D F T 部の出力と前記第 2 の D F T 部の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正する特性補正部と、

前記 1 入力スペクトルサブトラクション部の出力から前記特性補正部の出力を差し引くことにより 2 入力のスペクトルサブトラクションを行なう 2 入力スペクトルサブトラクション部を備え、前記 2 入力スペクトルサブトラクション部の出力を用いて音声認識を行なうことを特徴とする電話音声認識装置。

【請求項 10】電話の受信信号をスペクトルに変換する第 1 の D F T 部と、

電話の送信信号をスペクトルに変換する第 2 の D F T 部と、

前記第 1 の D F T 部の出力から相手の話者の音声の周囲雑音を除去する 1 入力スペクトルサブトラクション部と、

前記 1 入力スペクトルサブトラクション部の出力と前記第 2 の D F T 部の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、前記第 2 の D F T 部の出力信号を受信側の特性に合わせて補正する特性補正部

と、

前記 1 入力スペクトルサブトラクション部の出力から前記特性補正部の出力を差し引くことにより 2 入力のスペクトルサブトラクションを行なう 2 入力スペクトルサブトラクション部を備え、前記 2 入力スペクトルサブトラクション部の出力を用いて音声認識を行なうことを特徴とする電話音声認識装置。

【請求項 11】前記 2 入力スペクトルサブトラクション部の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部を備え、

前記 1 入力スペクトルサブトラクション部は前記音声検出部が音声区間検出中でない区間から除去する雑音を求めることを特徴とする請求項 9 または 10 記載の電話音声認識装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は電話回線の音声認識に用いられる音声除去装置に関し、特に相手の話者の周囲雑音と送信信号の受信側への回り込みとの両方を除去する技術に関する。

【0002】

【従来の技術】電話回線を用いた音声認識を行なう際には通常ガイダンス音声を用いて対話的なシステムを構成する場合が多い。例えば、以下にこの種の対話の例を示すと、

（ガイダンス）こちらは、〇〇商事でございます。御用件を発声して下さい。

（相手の話者）在庫照会

（ガイダンス）在庫照会でございますね。はいまたははいえでお答え下さい。

（相手の話者）はい

この対話の例において、相手の話者がある程度システムに習熟してくると、ガイダンスの終了を待たずに発声を行なう場合がある。例えば、第 1 番目のガイダンスにおいて「こちらは、〇〇商事でございます。」の直後に相手の話者が「在庫照会」の発声を行なう場合や第 2 番目の「在庫照会でございますね。」の直後に相手の話者が「はい」を発声する場合が考えられる。従来、音声認識装置はシステムが出力したガイダンスと相手の話者の音声とが同時に行なわれた場合、認識装置に入力される音声は 2 線-4 線変換器（ハイブリッド）で両者が混合されたものとなる。このような 2 つの音声为重畳する場合には正しく認識が行なえないため、このような発声を受理することが出来なかった。したがって、ある程度習熟した利用者においても、ガイダンスが終了するまで必ず待たなければならず、極めて冗長な、使い勝手の悪いシステムとなり、ひいてはシステムが利用されなくなる要因ともなっていた。

【0003】この種の音声除去装置はこのような問題点に対処するものであり、ガイダンス音声と相手の話者の

音声とが重畳した信号から、相手の話者の音声のみを抽出して音声認識を行なうことにより、利用者がガイダンスの終了を待たずに発声した音声を高精度に認識すること（いわゆるバージン）を可能とするものである。例えば、辻井（監修）：「エコーキャンセラ技術」、日本工業技術センター、p. 4（昭和 61 年 12 月）、（以下引用文献〔1〕とする）に示される音声除去装置を、図 10 を用いて説明する。図 10 に示した音声除去装置はいわゆるエコーキャンセラと呼ばれる技術を長距離電話回線の中継用交換機内部のハイブリッドにおける回り込み除去に応用した例であり、時間波形上の適応フィルタを用いて送信側の音声を受信側の音声に等価せしめ、時間波形上で引き算を行なうものである。ここでは、引用文献〔1〕記載の音声除去装置を加入者側の電話器に用いた音声認識装置に応用した場合について説明する。すなわち、ガイダンス音声が出力されるとハイブリッド 101 で漏洩し、4 線部分の受話側へ回り込む。疑似エコー発生部 103 は、残留エコーが最小となるように回り込み音声の伝送特性を補正し、疑似エコーを発生する。減算器 102 は、回り込みのガイダンス音声から疑似エコー発生部 103 が出力した疑似エコーを引き算することにより、ガイダンス音声のみをキャンセルする。電話回線から伝送された相手話者の音声とガイダンス音声とが重なった状態では、疑似エコー発生部 103 は最も直前に推定した伝送特性を利用して疑似エコーを発生し、減算器 102 はガイダンスに関するエコーのみを除去する。音声認識部 104 は、減算器 102 が処理した、ガイダンス音声除去後の音声を用いて音声認識を行なう。

【0004】図 10 に示す例は、時間波形上の適応フィルタを用いた例を示したが、これをパワースペクトル上の引き算（いわゆる 2 入力のスペクトルサブトラクション）で実現することも可能である。例えば、有吉、松下、藤本：「2 入力による騒音下の音声認識方式」、音響学会秋季講演論文集、1-8-5、pp. 9-10（1990. 9）、（以下引用文献〔2〕とする）に用いられている 2 入力スペクトルサブトラクションを用いれば、図 10 と同様の構成をパワースペクトル上の適応フィルタと引き算で実現することも可能である。なお、引用文献〔2〕では、その応用として自動車内の音声認識を行なった場合を示しており、この場合の除去対象は自動車内の周囲雑音となっているが、原理的には電話における音声除去技術と同一である。この方式は引用文献〔1〕記載の音声除去装置に比べて、パワースペクトル上で位相を除去した後に演算を行なうことで処理量を大きく低減化が可能であり、より廉価なシステムを構成することが可能である。

【0005】

【発明が解決しようとする課題】しかしながら、従来の音声除去装置では、除去する対象が自己の送出したガイ

ダンス音声のみであり、相手話者の周囲の雑音レベルが高い場合にもこの雑音に対する対処がなされておらず、音声認識性能が劣化するという問題点があった。

【0006】また、相手話者の周囲の雑音レベルが高い場合には、その影響により電話の送信信号の受話器側への回り込みに関する伝送特性が正しく推定出来ず、ひいてはガイダンス音声の除去性能が低下するという問題点があった。

【0007】本発明は上述の問題を解決するものであり、その目的は、自己の送出したガイダンス音声だけでなく相手話者の周囲の雑音をも除去し、高性能な音声および雑音の除去装置と、音声認識装置を提供することにある。

【0008】また、他の目的として相手話者の周囲の雑音影響をとり除くことで、自己の送出したガイダンス音声を高精度に除去することにある。

【0009】

【課題を解決するための手段】本発明による第1の音声および雑音の除去装置は、電話の受信信号をスペクトルに変換する第1のDFT部(図1の11)と、電話の送信信号をスペクトルに変換する第2のDFT部(図1の12)と、第1のDFT部(図1の11)の出力から相手の話者の音声の周囲雑音を除去する1入力スペクトルサブトラクション部(図1の13)と、第1のDFT部(図1の11)の出力と第2のDFT部(図1の12)の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、第2のDFT部(図1の12)の出力信号を受信側の特性に合わせて補正する特性補正部(図1の14)と、1入力スペクトルサブトラクション部(図1の13)の出力から特性補正部(図1の14)の出力を差し引くことにより2入力のスペクトルサブトラクションを行なう2入力スペクトルサブトラクション部(図1の15)とで構成される。

【0010】本発明による第2の音声および雑音の除去装置は、電話の受信信号をスペクトルに変換する第1のDFT部(図2の21)と、電話の送信信号をスペクトルに変換する第2のDFT部(図2の22)と、第1のDFT部(図2の21)の出力から相手の話者の音声の周囲雑音を除去する1入力スペクトルサブトラクション部(図2の23)と、1入力スペクトルサブトラクション部(図2の23)の出力と第2のDFT部(図2の22)の出力とから送信信号の受信側への回り込みに関する伝送特性を推定し、第2のDFT部(図2の22)の出力信号を受信側の特性に合わせて補正する特性補正部(図2の24)と、1入力スペクトルサブトラクション部(図2の23)の出力から特性補正部(図2の24)の出力を差し引くことにより2入力のスペクトルサブトラクションを行なう2入力スペクトルサブトラクション部(図2の25)とで構成される。

【0011】本発明による第3の音声および雑音の除去

装置は、本発明による第1または第2の音声および雑音の除去装置の特性補正部(図3の34)が送信信号の存在する区間を用いて電話の送信信号の受信側への回り込みに関する伝送特性を推定し、DFT部2(図3の32)の出力信号を受信側の特性に合わせて補正する。

【0012】本発明による第4の音声および雑音の除去装置は、本発明による第1または第2の音声および雑音の除去装置に加えて2入力スペクトルサブトラクション部(図4の45)の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部(図4の46)を備え、1入力スペクトルサブトラクション部(図4の43)は音声検出部(図4の46)が音声区間検出中でない区間から除去する雑音を求めることを特徴とする。

【0013】本発明による第5の音声および雑音の除去装置は、本発明による第4の音声および雑音の除去装置の1入力スペクトルサブトラクション部(図5の53)が音声検出部(図5の56)が音声区間検出中でない区間かつ電話の送信信号の存在しない区間から除去する雑音を求めることを特徴とする。

【0014】本発明による第6の音声および雑音の除去装置は、本発明による第4または第5の音声および雑音の除去装置の特性補正部(図6の64)が音声検出部(図6の66)が音声区間検出中でない区間から電話の送信信号の受話器側への回り込みに関する伝送特性を推定し、第2のDFT部(図6の62)の出力信号を受信側の特性に合わせて補正することを特徴とする。

【0015】本発明による第7の音声および雑音の除去装置は、本発明による第6の音声および雑音の除去装置の特性補正部(図7の74)が音声検出部(図7の76)が音声区間検出中でない区間かつ電話の送信信号の存在する区間から電話の送信信号の受話器側への回り込みに関する伝送特性を推定し、第2のDFT部(図7の72)の出力信号を受信側の特性に合わせて補正することを特徴とする。

【0016】本発明による第8の音声および雑音の除去装置は、本発明による第3の音声および雑音の除去装置の2入力スペクトルサブトラクション部(図8の85)の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部(図8の86)を備え、特性補正部(図8の84)が音声検出部(図8の86)が音声区間検出中でない区間かつ電話の送信信号の存在する区間を用いて送信信号の受信側への回り込みに関する伝送特性を推定し、第2のDFT部(図8の82)の出力信号を受信側の特性に合わせて補正することを特徴とする。

【0017】本発明による音声認識装置は、本発明の音声および雑音の除去装置の出力を用いて音声認識を行なう音声認識部(図9の91)を有する。

【0018】本発明は、電話における相手話者の音声と送信したガイダンス音声とが重畳した信号から相手話者の音声のみを取りだし、しかも相手話者の周囲雑音を除

去することにより、高性能な雑音および音声の除去装置を提供出来るという効果を得るものであり、従来の音声の除去装置では得られなかった高い雑音および音声の除去性能を得るものである。

【0019】本発明の第2の音声および雑音の除去装置の作用を図2を用いて説明する。まず、ガイダンスとして送出した信号はハイブリッドを経て電話回線に送出されるのと同時にハイブリッドの受信信号線にも漏洩す

$$x(t) = s(t) + n(t)$$

送出するガイダンス信号を $y(t)$ 、ハイブリッドの回り込みのインパルス応答特性を $h(t)$ とすると、ハイブリッドの受信側信号線の観測信号 $w(t)$ は以下のよ

$$\begin{aligned} w(t) &= x(t) + h(t) \otimes y(t) \\ &= s(t) + n(t) + h(t) \otimes y(t) \end{aligned}$$

【0021】なお、

【0022】

【外1】

⊗

【0023】は波形同士の畳み込み演算を表すものとする。また、実際は相手話者の音声と周囲雑音との合成信号 $x(t)$ 自体もハイブリッドの伝達特性により変形を

$$W(f) = S(f) + N(f) + H(f)Y(f)$$

$$V(f) = Y(f)$$

の関係が得られる。ここで、 $S(f)$ 、 $N(f)$ 、 $H(f)$ 、 $Y(f)$ はそれぞれ、 $s(t)$ 、 $n(t)$ 、 $h(t)$ 、 $y(t)$ のパワースペクトルを表すものとする。

【0024】いま、相手話者の周囲雑音 $N(f)$ が定常であると仮定すれば、 $N(f)$ は、相手話者の音声 $S(f)$ およびガイダンス音声 $Y(f)$ がどちらも存在しない部分で直接観測することが可能である。例えば、相手話者と回線が接続された瞬間かつガイダンス音声を出力していない時間位置で観測することが出来る。また、

$$\begin{aligned} W_1(f) &= W(f) - N'(f) \\ &= S(f) + N(f) + H(f)Y(f) - N'(f) \\ &\approx S(f) + H(f)Y(f) \end{aligned} \quad (4)$$

【0026】の様に話者の周囲雑音が除去されることになる。特性補正部(図2の24)は、まずDFT部2(図2の22)と1入カスペクトルサブトラクション部(図2の23)とから、ハイブリッドのパワー伝送特性 $H(f)$ を推定する。いま相手話者の音声が存在せずが

$$W_1(f) \approx H(f)Y(f) \quad (S(f)=0)$$

【0028】のようにハイブリッドで変形を受けたガイダンス音声のみとなり、この時間位置を用いて伝送特性 $H'(f)$ を事前に推定することが出来る。

$$\begin{aligned} H'(f) &= W_1(f)/Y(f) \\ &\approx H(f)Y(f)/Y(f) \quad (S(f)=0) \\ &= H(f) \end{aligned}$$

る。相手話者がガイダンス音声と時間的に重なった位置で発声を行なった場合は、受信信号線には両方の音声为重畳した信号が観測される。相手話者の音声に周囲雑音が存在する場合は、受信信号にはさらに相手話者の周囲雑音が重畳される。相手話者の音声を $s(t)$ 、周囲雑音を $n(t)$ とし、相手話者の周囲雑音が重畳された観測信号を $x(t)$ とすると、

(1)

うに表される。

【0020】

【数1】

(2)

受けるが、本発明の目的とは直接関係がないので考慮に入れないものとする。DFT部1(図2の21)およびDFT部2(図2の22)は、それぞれハイブリッドの送信信号および受信信号をパワースペクトルに変換する。各々の出力のパワースペクトルを $W(f)$ 、 $V(f)$ とすると、

(3)

音声検出手段を設けて観測することももちろん可能である。1入カスペクトルサブトラクション部(図2の23)は、この様にして求めた周囲雑音 $N'(f)$ をDFT部1(図2の21)の出力スペクトルから引き算することにより話者の周囲雑音の除去を行なう。1入カスペクトルサブトラクション(図2の23)の出力スペクトルを $W_1(f)$ とすると、

【0025】

【数2】

イダンス音声のみが存在する場合、1入カスペクトルサブトラクション部(図2の23)の出力は、

【0027】

【数3】

(5)

【0029】

【数4】

(6)

【0030】特性補正部（図2の24）はこの伝送特性  $H'(f)$  を DFT 部 2（図2の22）の出力に掛けることにより出力  $V'(f)$  を得る。2入力スペクトルサブトラクション部（図2の25）は、1入力スペクトルサブトラクション部（図2の23）の出力  $W_1(f)$  が

$$\begin{aligned} W_2(f) &= W_1(f) - V'(f) \\ &= S(f) + H(f)Y(f) - H'(f)Y(f) \\ &\approx S(f) \end{aligned}$$

【0032】以上のように、2入力のスペクトルサブトラクションを行なう前に相手話者の周囲雑音を除去する1入力スペクトルサブトラクション部（図2の23）を設けることで、相手話者の周囲雑音が多い場合にも最終的に得られたスペクトル  $W_2(f)$  は相手話者の周囲雑音およびガイダンス音声の両方が除去されており、高精度な音声および雑音の除去装置の提供が可能である。

【0033】なお、1入力スペクトルサブトラクション部を設けない従来の音声除去装置においては、2入力のスペクトルサブトラクション部が出力するスペクトルに相手話者の周囲雑音および除去できなかったガイダンス音声が入混することはここで述べた方法により容易に導出可能であることはいうまでもない。

【0034】

【発明の実施の形態】以下、図1から図9を参照しながら本発明の実施の形態について説明する。

【0035】図1は本発明の第1の音声および雑音の除去装置の実施の形態のブロック図である。この音声および雑音の除去装置は、電話の受信信号をスペクトルに変換する第1の DFT 部 11 と、ガイダンス等の電話の送

$$x(t) = s(t) + n(t)$$

送出するガイダンス信号を  $y(t)$ 、ハイブリッドの回り込みのインパルス応答特性を  $h(t)$  とすると、ハイブリッドの受信側信号線の観測信号  $w(t)$  は以下のよ

$$\begin{aligned} w(t) &= z(t) + h(t) \otimes v(t) \\ &= s(t) + n(t) + h(t) \otimes v(t) \end{aligned}$$

【0038】なお、

【0039】

【外2】

⊗

【0040】は波形同士の畳み込み演算を表すものとする。また、実際は相手話者の音声と周囲雑音との合成信号  $x(t)$  自体もハイブリッドの伝達特性により変形を受けるが、本発明の目的とは直接関係がないので考慮に

$$W(f) = S(f) + N(f) + H(f)Y(f)$$

$$V(f) = Y(f)$$

の関係が得られる。ここで、 $S(f)$ 、 $N(f)$ 、 $H(f)$ 、 $Y(f)$  はそれぞれ、 $s(t)$ 、 $n(t)$ 、 $h(t)$ 、 $y(t)$  のパワースペクトルを表すものとする。いま、相手話者の周囲雑音  $N(f)$  が定常であると仮定すれば、 $N(f)$  は、相手話者の音声  $S(f)$  およびガイダンス音声  $Y(f)$  がどちらも存在しない部分で

ら特性補正部（図2の24）の出力  $V'(f)$  を引くことにより、 $W_2(f)$  を出力する。

【0031】

【数5】

(7)

信信号をスペクトルに変換する第2の DFT 部 12 と、第1の DFT 部 11 の出力から相手の話者の音声の周囲雑音を除去する1入力スペクトルサブトラクション部 13 と、第1の DFT 部 11 の出力と第2の DFT 部 12 の出力とから電話の送信信号の受信側への回り込みに関する伝送特性を推定し、第2の DFT 部 12 の出力信号を受信側の特性に合わせて補正する特性補正部 14 と、1入力スペクトルサブトラクション部 13 の出力から特性補正部 14 の出力を差し引くことにより2入力のスペクトルサブトラクションを行なう2入力スペクトルサブトラクション部 15 とで構成される。

【0036】ガイダンスとして送出した信号はハイブリッドを経て電話回線に送出されるのと同時にハイブリッドの受信信号線にも漏洩する。相手話者がガイダンス音声と時間的に重なった位置で発声を行なった場合は、受信信号線には両方の音声を重ねた信号が観測される。相手話者の音声に周囲雑音が存在する場合は、受信信号にはさらに相手話者の周囲雑音を重ねられる。相手話者の音声を  $s(t)$ 、周囲雑音を  $n(t)$  とし、相手話者の周囲雑音を重ねられた観測信号を  $x(t)$  とすると、

(8)

うに表される。

【0037】

【数6】

(9)

入れないものとする。第1の DFT 部 11 および第2の DFT 部 12 は、それぞれハイブリッドの送信信号および受信信号をパワースペクトルに変換する。ここでは、デジタル信号に変換した後に DFT（離散的フーリエ変換）を行なう例を用いたが、他のスペクトル変換手法や、フィルタバンクなどを利用することも可能である。第1の DFT 部 11 および第2の DFT 部 12 各々の出力のパワースペクトルを  $W(f)$ 、 $V(f)$  とすると、

(10)

直接観測することが可能である。例えば、相手話者と回線が接続された瞬間かつガイダンス音声を出力していない時間位置で観測することが出来る。また、1入力スペクトルサブトラクション部 13 は、この様にして求めた周囲雑音  $N'(f)$  を第1の DFT 部 11 の出力スペクトルから引き算することにより話者の周囲雑音の除去を

行なう。1入力スペクトルサブトラクションは従来より種々の方式が公知例として知られているので、ここでは詳細な説明は省略する。例えば、S. F. Boll:

"Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. on ASSP., Vol. ASSP-27,

$$\begin{aligned} W_1(f) &= W(f) - N'(f) \\ &= S(f) + N(f) + H(f)Y(f) - N'(f) \\ &\approx S(f) + H(f)Y(f) \end{aligned} \quad (11)$$

【0042】の様に話者の周囲雑音が除去されることになる。引き算の結果スペクトルが負になった成分は0としたり、あるいは引き算の結果の下限を設けてそれ以下の値を取らないような工夫が引用文献【3】で述べられており、この様な工夫を設けることももちろん可能であり、この他1入力のスペクトルサブトラクションの原理を応用した方式であればどのような方法も利用することが可能である。また、推定した雑音 $N'(f)$ は、比較的 $W(f) \approx H(f)Y(f)$  ( $N(f) = S(f) = 0$ )

【0044】のようにハイブリッドで変形を受けたガイダンス音声のみとなり、この時間位置あるいはその区間の平均値を用いて伝送特性 $H'(f)$ を事前に推定する

$$\begin{aligned} H'(f) &= W(f)/V(f) \\ &\approx H(f)Y(f)/Y(f) \quad (N(f) = S(f) = 0) \\ &= H(f) \end{aligned} \quad (13)$$

【0046】特性補正部14はこの伝送特性 $H'(f)$ を第2のDFT部12の出力に掛けることにより出力 $V'(f)$ を得る。2入力スペクトルサブトラクション部15は、1入力スペクトルサブトラクション部13の

$$\begin{aligned} W_2(f) &= W_1(f) - V'(f) \\ &= S(f) + H(f)Y(f) - H'(f)Y(f) \\ &\approx S(f) \end{aligned} \quad (14)$$

【0048】2入力スペクトルサブトラクションについても従来より種々の方法が公知となっているが、引用文献【2】で挙げた方法の他、2入力スペクトルサブトラクションの原理を応用したものであればどのような方法も利用可能である。

【0049】以上のように、2入力のスペクトルサブトラクションを行なう前に相手話者の周囲雑音を除去する1入力スペクトルサブトラクション部13を設けることで、相手話者の周囲雑音が大きい場合にも最終的に得られたスペクトル $W_2(f)$ は相手話者の周囲雑音およびガイダンス音声の両方が除去されており、高精度な音声および雑音の除去装置の提供が可能である。

【0050】図2は本発明の第2の音声および雑音の除去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、電話の受信信号をスペクトルに変換する第1のDFT部21と、ガイダンス等の電話の送信信号をスペクトルに変換する第2のDFT部22

No. 2, pp. 113-120 (1979, 4).

(以下引用文献【3】とする)などがその一例である。1入力スペクトルサブトラクション部13の出力スペクトルを $W_1(f)$ とすると、

【0041】

【数7】

長時間の平均値を用いることも可能である。特性補正部14は、まず第2のDFT部12と第1のDFT部11とから、ハイブリッドのパワー伝送特性 $H(f)$ を推定する。いま相手話者の音声および周囲雑音が存在せずガイダンス音声のみが存在する場合、第1のDFT部11の出力は、

【0043】

【数8】

(12)

ことが出来る。

【0045】

【数9】

出力 $W_1(f)$ から特性補正部14の出力 $V'(f)$ を引くことにより、 $W_2(f)$ を出力する。

【0047】

【数10】

と、第1のDFT部21の出力から相手の話者の音声の周囲雑音を除去する1入力スペクトルサブトラクション部23と、1入力スペクトルサブトラクション部23の出力と第2のDFT部22の出力とから電話の送信信号の受信側への回り込みに関する伝送特性を推定し、第2のDFT部22の出力信号を受信側の特性に合わせて補正する特性補正部24と、1入力スペクトルサブトラクション部23の出力から特性補正部24の出力を差し引くことにより2入力のスペクトルサブトラクションを行なう2入力スペクトルサブトラクション部25とで構成される。

【0051】図2の音声および雑音の除去装置は基本的に図1の音声および雑音の除去装置と同じ動作をするが、その違いは、図2の特性補正部24が、第1のDFT部21の出力を用いるのではなく、1入力スペクトルサブトラクション部23と第2のDFT部22の出力とから回り込みに関する伝送特性を推定する点にある。こ

れは、周囲雑音が存在しない場合を仮定した図 1 の方法より相手話者の音声のみ存在しない場合を仮定した図 2 の場合の方が特性推定に利用可能な区間候補が長く取れる場合があるという理由による。すなわち、特性補正部 24 は、相手話者の周囲雑音のみ存在する状況において

$$W_1(f) \approx H(f)Y(f) \quad (S(f)=0)$$

【0053】となり、この時間位置を用いて伝送特性  $H'(f)$  を事前に推定することが出来る。

$$\begin{aligned} H'(f) &= W_1(f)/V(f) \\ &\approx H(f)Y(f)/Y(f) \quad (N(f)=S(f)=0) \\ &= H(f) \end{aligned}$$

【0055】以下、図 1 と同様の方法により、最終的に 2 入力スペクトルサブトラクション部 25 の出力は相手話者の周囲雑音およびガイダンス音声の両方が除去されたものとなる。

【0056】図 3 は本発明の第 3 の音声および雑音の除去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、本発明の第 1 または第 2 の音声および雑音の除去装置の特性補正部 34 が、電話の送信信号の存在する区間を用いて送信信号の受信側への回り込みに関する伝送特性を推定し、第 2 の DFT 部 32 の出力信号を受信側の特性に合わせて補正する。なお、図 3 は図 1 に示す本発明の第 1 の音声および雑音の除去装置に対して実施した場合について示してある。

【0057】電話の送信信号の存在する区間の同定については種々の方法が考えられる。図 3 にはガイダンス区間信号を外から供給し、その信号を利用した場合についてののみ示してあるが、その他第 2 の DFT 部 32 の出力を直接利用してガイダンス音声区間を求めたり、あるいは簡易なパワー判定を利用してガイダンス信号そのものを用いることが可能である。なお、従来より音声の検出方法については種々のものが公知となっており、音声の存在する区間を同定出来るものであればどのような方法を用いても良い。ガイダンス音声は通常相手話者との対話のなかで開始、終了、中断が頻繁に繰り返されるため、ガイダンス音声が存在しない区間がある。このような場合、ガイダンス音声のない部位の信号を用いたり、あるいはガイダンス音声のない部位の信号を含めて平均を求めたりすると伝送特性  $H'(f)$  の推定精度が低下する虞がある。図 3 の本発明の第 3 の音声および雑音の除去装置はこのような問題点に対処したものである。

【0058】図 4 は本発明の第 4 の音声および雑音の除去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、本発明の第 1 または第 2 の音声および雑音の除去装置の 2 入力スペクトルサブトラクション部 45 の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部 46 を有し、1 入力スペクトルサブトラクション部 43 が音声検出部 46 が音声区間検出中でない区間から除去する雑音を求める構成

て、ガイダンス音声のみが存在する場合、1 入力スペクトルサブトラクション部 13 の出力は、

【0052】

【数 11】

(15)

【0054】

【数 12】

(16)

となっている。なお、図 4 は図 1 の音声および雑音の除去装置に対して実施した場合について示してある。

【0059】1 入力スペクトルサブトラクション部 43 は、相手の話者の周囲雑音のみ存在する区間を用いて周囲雑音のスペクトルを推定する。例えば、簡便な方法として電話回線が接続された直後の部位を用いて周囲雑音のスペクトルを推定することが可能であるが、相手話者がすぐに発声を始めた場合などには推定したスペクトルに音声が入り込み 1 入力スペクトルサブトラクションの雑音除去効果が低下する場合がある。このような場合においても安定に動作する装置の提供が可能である。音声検出部 46 は、2 入力スペクトルサブトラクション部 45 が出力したスペクトル情報を用いて相手話者の音声の開始、終了を検出する。なお、従来より音声の検出方法については種々のものが公知となっており、音声の存在する区間を同定出来るものであればどのような方法を用いても良い。なお、音声検出部 46 は、最終的に求めたスペクトルのみ出力しても良いが、検出した音声の始端終端情報を出力したり、あるいは検出した音声の始端から終端までのスペクトルのみ出力するような構成も可能である。

【0060】図 5 は本発明の第 5 の音声および雑音の除去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、本発明の第 4 の音声および雑音の除去装置に加えて、1 入力スペクトルサブトラクション部 53 は音声検出部 56 が音声区間検出中でない区間かつ電話の送信信号の存在しない区間から除去する雑音を求める。

【0061】1 入力スペクトルサブトラクション部 53 は、相手の話者の周囲雑音のみ存在する区間かつガイダンス音声が存在しない区間を用いて周囲雑音のスペクトルを推定する。このような構成により、図 4 の本発明の第 4 の音声および雑音の除去装置の効果に加えてガイダンス音声のない部位の信号を用いたり、あるいはガイダンス音声のない部位の信号を含めて平均を求めたりすると求めた推定雑音にガイダンス音声が入り込み雑音の推定精度が低下する虞があるものを解決するものである。

【0062】図 6 は本発明の第 6 の音声および雑音の除

去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、本発明の第4または第5の音声および雑音の除去装置に加えて、特性補正部64が音声検出部66が音声区間検出中でない区間から電話の送信信号の受信側側への回り込みに関する伝送特性を推定し、第2のDFT部62の出力信号を受信側の特性に合わせて補正する。

【0063】このような構成となすことにより、本発明の第4または第5の音声および雑音の除去装置が有している効果に加えて、特性補正部64の回り込みに関する特性の推定時に相手話者の音声が入混入して推定精度が低下することを防ぎ、ひいては高い除去性能を得ることが可能であるという効果を有する。

【0064】図7は本発明の第7の音声および雑音の除去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、本発明の第6の音声および雑音の除去装置に加えて、特性補正部74が音声検出部76が音声区間検出中でない区間かつ電話の送信信号の存在する区間から電話の送信信号の受信側側への回り込みに関する伝送特性を推定し、第2のDFT部72の出力信号を受信側の特性に合わせて補正する。

【0065】このような構成となすことにより、本発明の第6の音声および雑音の除去装置が有している効果に加えて、ガイダンスが確実に存在する区間のみを用いて特性推定を行なうことが出来、ひいては高い除去性能を得ることが可能であるという効果を有する。

【0066】図8は本発明の第8の音声および雑音の除去装置の実施の形態を示すブロック図である。この音声および雑音の除去装置は、本発明の第3の音声および雑音の除去装置に加えて、2入力スペクトルサブトラクション部85の出力を用いて相手の話者の音声の存在する区間の検出を行なう音声検出部86を備え、特性補正部84が音声検出部86が音声区間検出中でない区間かつ電話の送信信号の存在する区間を用いて送信信号の受信側側への回り込みに関する伝送特性を推定し、第2のDFT部82の出力信号を受信側の特性に合わせて補正する。

【0067】このような構成となすことにより、本発明の第3の音声および雑音の除去装置が有している効果に加えて、ガイダンスが確実に存在し、しかも相手話者の音声が存在しない区間で回り込みに関する伝送特性を推定するため、推定の精度が高く、ひいては高い除去性能を得ることが可能であるという効果を有する。なお、音声検出部86は、最終的に求めたスペクトルのみ出力しても良いが、検出した音声の始端・終端情報を出力したり、あるいは検出した音声の始端から終端までのスペクトルのみ出力するような構成も可能である。

【0068】図9は本発明の音声および雑音の除去装置を用いた音声認識装置の実施の形態を示すブロック図である。前記音声および雑音の除去装置の出力を用いて音

声認識を行なう音声認識部91を有する。

【0069】認識部91は本発明の音声および雑音の除去装置が出力するスペクトルを元に音声認識のための特徴ベクトルの時系列に変換し、音声認識を行ない、認識結果を出力する。認識部91は通常音声検出手段を有しているが、本発明の第4～第8の音声および雑音の除去装置においては音声検出部46、56、66、76、86を有しており、この場合にはこれらの検出情報を利用したり、あるいは音声検出部46、56、66、76、86が切り出した音声区間のスペクトルのみを利用するような構成にすることも可能である。このような構成により、本発明の音声および雑音の除去装置が有する高い雑音と音声の除去性能により高性能な音声認識装置が提供可能となる。

【0070】

【発明の効果】以上から明らかなように、本発明の音声および雑音の除去装置によれば、1入力のスペクトルサブトラクションを用いて事前に相手話者の周囲雑音を除去することで、相手の話者の周囲雑音をも除去可能な高精度な装置の提供が可能となった。

【0071】また、第2の効果としては、相手話者の周囲の雑音レベルが高い場合にも、その雑音に影響されず、回り込みに関する伝送特性が正しく推定出来るため、ガイダンス音声を高精度に除去可能であるという効果を有する。

【0072】第3の効果としては、ガイダンス音声の存在する区間のみを用いて回り込みに関する伝送特性を推定することで、より正確な推定伝送特性の推定が可能となり、高精度な装置の提供が可能となった。

【0073】第4の効果としては、音声検出部を設けその出力を利用し、相手話者の音声が存在しない区間で回り込みに関する伝送特性を推定することで、より正確な推定伝送特性の推定が可能となり、高精度な装置の提供が可能となった。

【0074】第5の効果としては、ガイダンス音声の存在しない区間のみを用いて1入力スペクトルサブトラクションのための周囲雑音の推定を行なうことで、より正確な周囲雑音の除去が可能となり、ひいては高精度な装置の提供が可能となった。

【0075】第6の効果としては、音声検出部を設けその出力を利用し、相手話者の音声が存在しない区間で1入力スペクトルサブトラクションのための周囲雑音の推定を行なうことで、より正確な周囲雑音の除去が可能となり、ひいては高精度な装置の提供が可能となった。

【0076】第7の効果としては、本発明の音声および雑音の除去装置を利用した音声認識装置は、その高精度な除去性能により従来にない高性能な音声認識装置の提供が可能となる。

【図面の簡単な説明】

【図1】本発明による音声および雑音の除去装置の一実

施の形態を示すブロック図である。

【図2】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

【図3】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

【図4】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

【図5】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

【図6】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

【図7】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

【図8】本発明による音声および雑音の除去装置の一実施の形態を示すブロック図である。

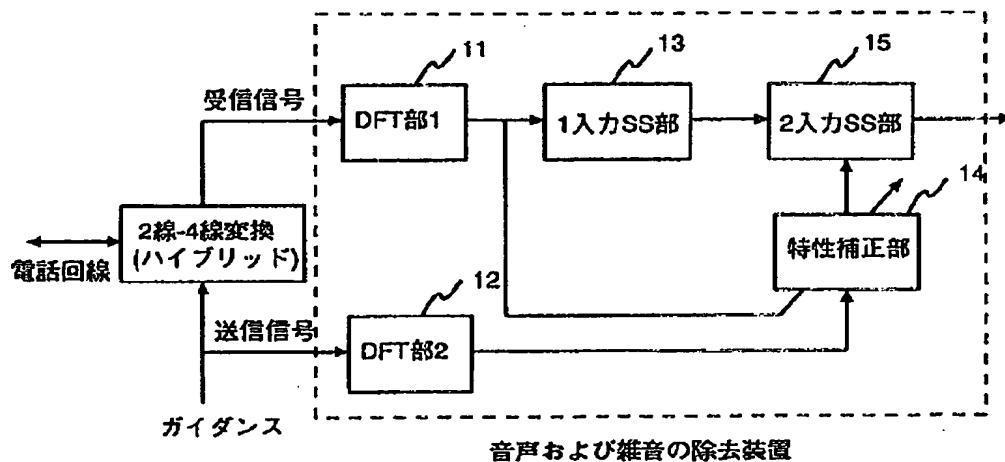
【図9】本発明による音声および雑音の除去装置を用いた音声認識装置の一実施の形態を示すブロック図である。

【図10】従来の音声の除去装置の一例を示すブロック図である。

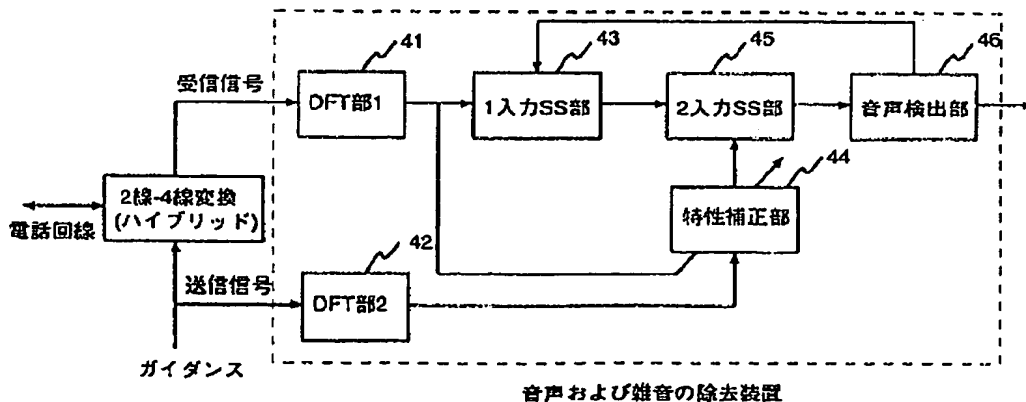
【符号の説明】

11, 21, 31, 41, 51, 61, 71, 81 第1のDFT部  
12, 22, 32, 42, 52, 62, 72, 82 第2のDFT部  
13, 23, 33, 43, 53, 63, 73, 83 1入力スペクトルサブトラクション部  
14, 24, 34, 44, 54, 64, 74, 84 特性補正部  
15, 25, 35, 45, 55, 65, 75, 85 2入力スペクトルサブトラクション部  
46, 56, 66, 76, 86 音声検出部  
91, 104 音声認識部  
101 ハイブリッド  
102 減算器  
103 疑似エコー発生部

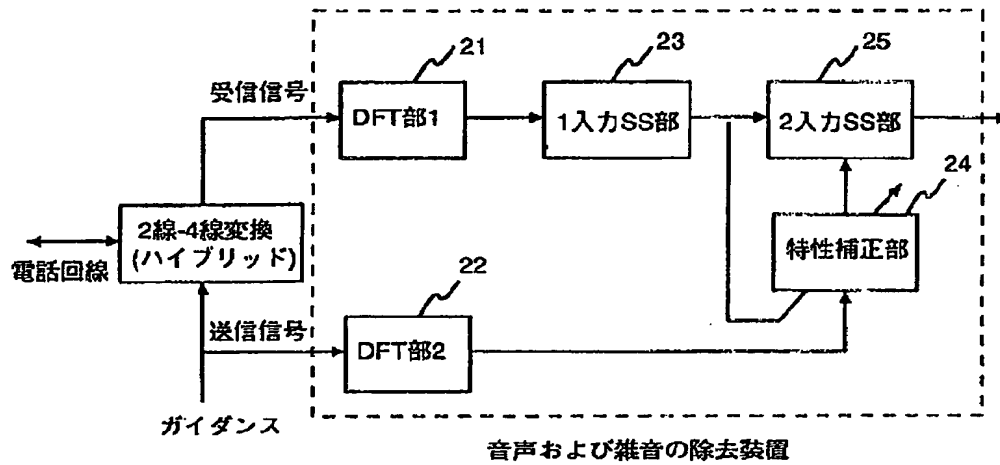
【図1】



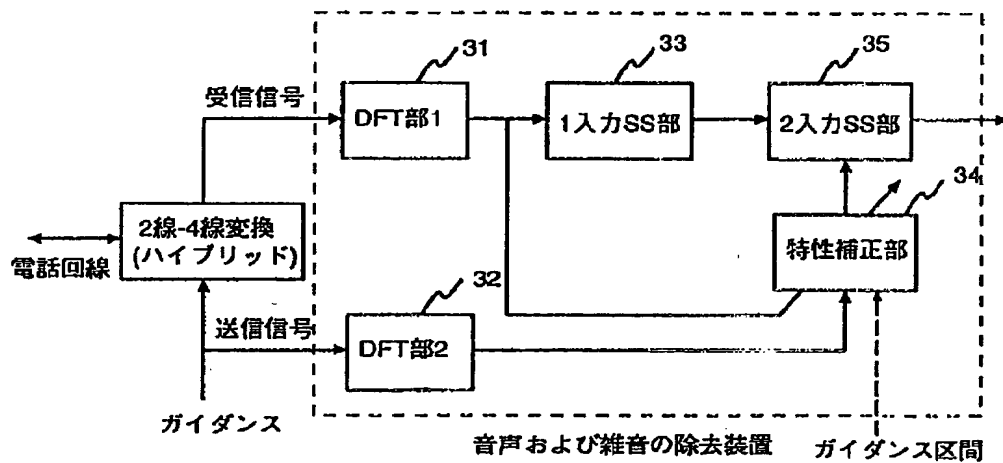
【図4】



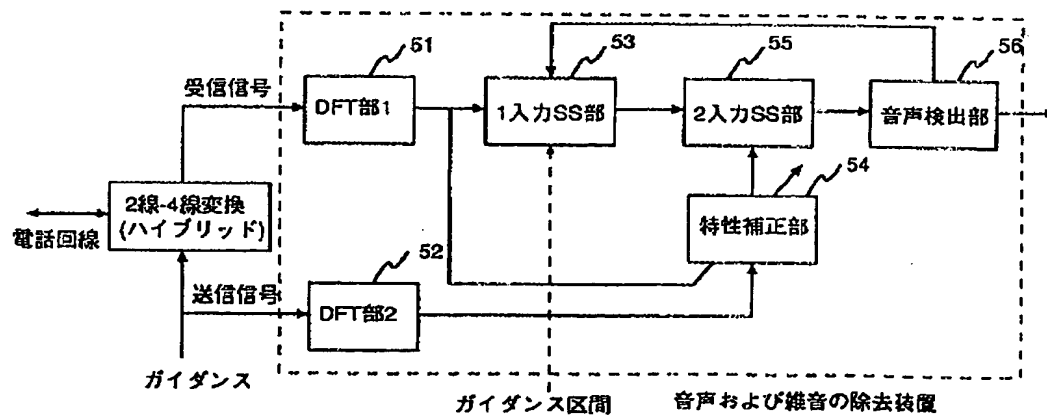
【図 2】



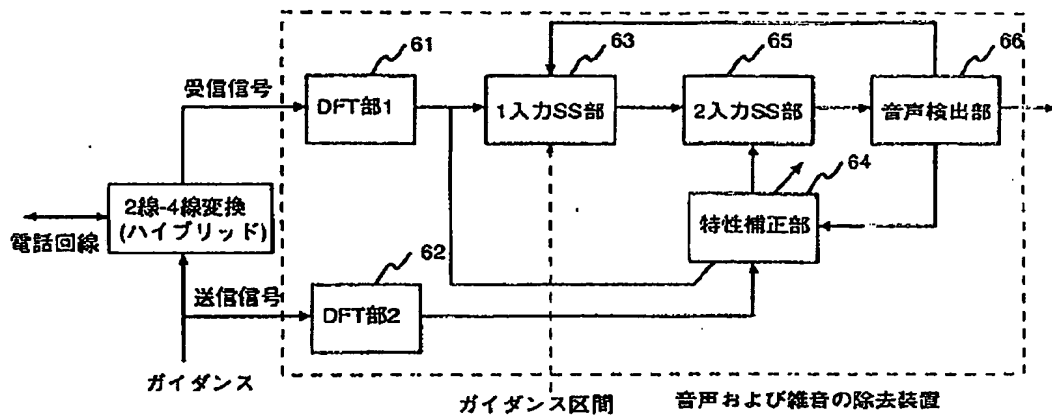
【図 3】



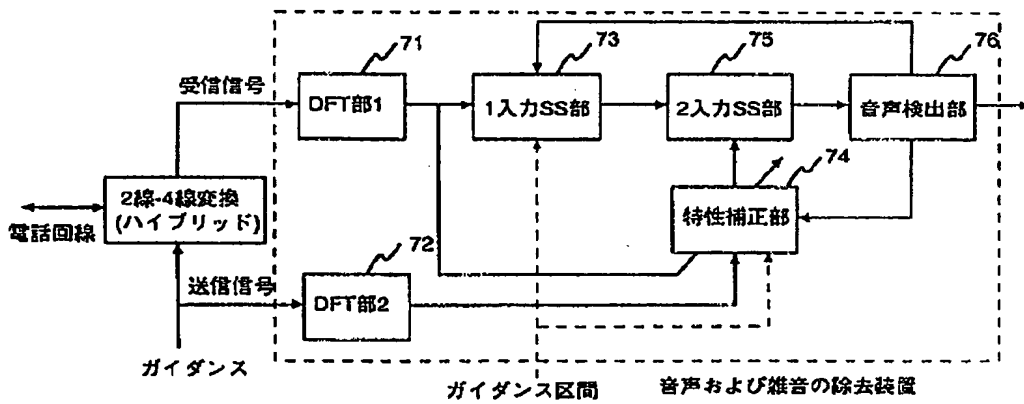
【図 5】



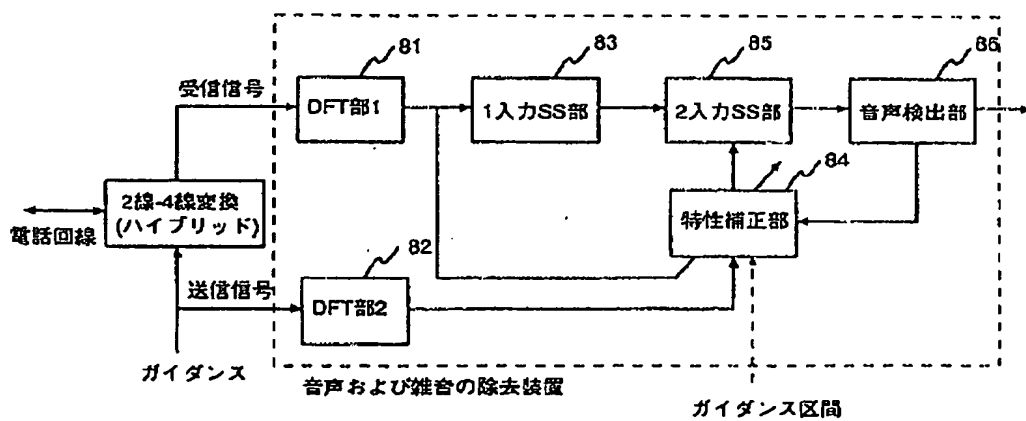
【図6】



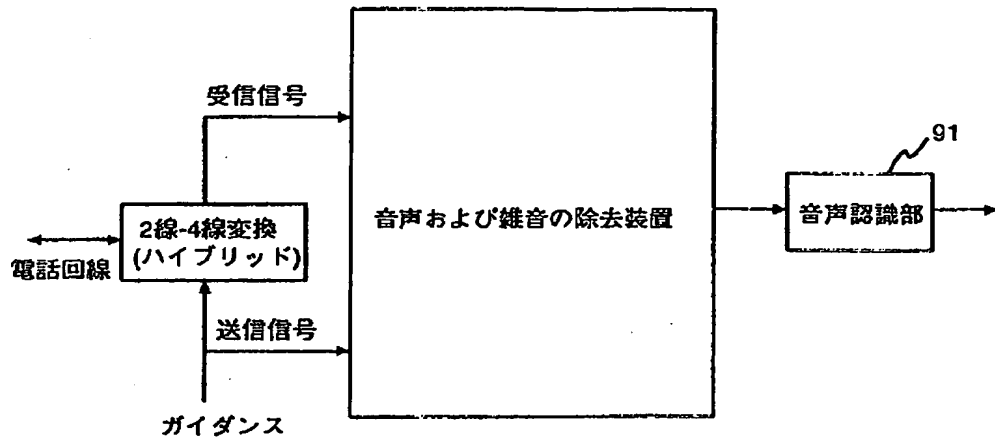
【図7】



【図8】



【図 9】



【図 10】

